

TARTU ÜLIKOOL

Loodus- ja täppisteaduste valdkond

Füüsika Instituut

Meri Liis Treimann

ÜHILDATUD GETM-ERGOM MUDELI  
VALIDEERIMINE SOOME LAHE  
TEMPERATUURI, NITRAADI JA  
KLOROFÜLLI PÕHJAL AASTATE  
1966–2005 KOHTA

Bakalaureusetöö

Juhendajad:

TTÜ MSI nooremteadur Ilja Maljutenko

TÜ FI dotsent Hanno Ohvri

Tartu 2016

# Sisukord

|  |           |
|--|-----------|
| <b>Sissejuhatus</b>                        | <b>4</b>  |
| <b>1 Mudelid</b>                           | <b>5</b>  |
| 1.1 GETM . . . . .                         | 5         |
| 1.2 ERGOM . . . . .                        | 5         |
| <b>2 Metoodika</b>                         | <b>8</b>  |
| 2.1 Andmed . . . . .                       | 8         |
| 2.2 Valideerimismeetodid . . . . .         | 12        |
| 2.2.1 Statistilised näitajad . . . . .     | 12        |
| 2.2.2 Graafiline võrdlus . . . . .         | 14        |
| 2.3 Arvutused . . . . .                    | 15        |
| <b>3 Tulemused ja arutelu</b>              | <b>17</b> |
| 3.1 Temperatuur . . . . .                  | 17        |
| 3.2 Nitraadis sisalduv lämmastik . . . . . | 20        |
| 3.3 Klorofüll . . . . .                    | 24        |
| 3.4 Arutelu . . . . .                      | 29        |
| <b>Kokkuvõte</b>                           | <b>30</b> |
| <b>Summary</b>                             | <b>32</b> |
| <b>Viited</b>                              | <b>34</b> |

## Sissejuhatus

Looduslikke süsteeme simuleerivate numbriliste mudelite õigsuse kinnitamine on võimatu, kuna looduslikud süsteemid ei ole kunagi kinnised. Samuti tehakse mudelites lähendusi, mistõttu kaob mõni väikesel skaalal toimuv protsess, mis võib pikaajalisel simulatsioonis suures mastaabis palju muuta. Samas kasutatakse mudeleid palju tulevikku mõjutavate otsuste tegemisel ning seetõttu on mingisugune hinnang mudeli sooritusele vajalik. Mudeli efektiivsust tegeliku maailma ennustamisel on võimalik hinnata –valideerida, võrreldes simulatsiooni väljundit vaatlusandmetega. Selleks võib kasutada nii statistilist analüüsi kui ka graafilist võrdlust.

Käesolevas töös võrreldakse Soome lahe füüsikalist ja biogeokeemilisi välju ühildatud GETM-ERGOM mudeliga ning teostatakse mudeli valideerimine, kasutades viie monitooringu jaama andmeid temperatuuri, nitraadi lämmastiksisalduse ja klorofüllü kontsentratsiooni kohta aastatel 1966–2005. Selleks kasutatakse kõrvuti nii statistikuid kui ka graafikuid, kuna ühelt poolt ei pruugi üks arv anda täielikku infot selle kohta, kui hästi ennustab mudel tegelikke väärtusi erinevates dimensioonides (aeg ja ruum) või tegeliku väärtuse suurusjärku. Teiselt poolt võib vaid graafiline võrdlus igale vaatajale mudelist erineva mulje jätta.

Töö on jaotatud kolmeks suuremaks peatükiks. Esimeses peatükis antakse lühike ülevaade mudelistest GETM ja ERGOM ning tuuakse asjakohased avaldised mudeli arvutatavate suuruste kohta. Teises peatükis kirjeldatakse esmalt jaamasid, milles kogutud andmete põhjal mudelit valideeritakse. Samuti kirjeldatakse andmetega seonduvat: nende kogumist, formaati ning arvutuste-eelset töötlemist. Lisaks tutvustatakse töös kasutatud nii statistilisi kui graafilisi valideerimismeetodeid ning kolmandaks kirjeldatakse andmete peal tehtud arvutusi. Kolmandas peatükis esitletakse lugejale tulemusi ning arutletakse kasutatud valideerimismeetodite ning tulemuste üle.

# 1 Mudelid

## 1.1 GETM

GETM (*General Estuarine Transport Model*) on kolmemõõtmeline numbriline mudel hüdrodünaamika simuleerimiseks, mis baseerub momendi jäävuse võrranditel. GETM loodi peamiselt ranniku- ning suudmealadel kasutamiseks, kuna arvutab erinevalt paljudest teistest hüdrodünaamika mudelitest ka vaba veepinna muutust, s.t veetaseme tõusu ja langust. [1]

GETMi põhjalikum kirjeldus on toodud artiklis [2].

Lisaks hoovuste dünaamikale arvutab GETM ka temperatuuri  $T$ , milles arvestatakse vee liikumist horisontaal- ja vertikaalsuunas ning leitakse valemiga

$$\partial_t T + \partial_x(uT) + \partial_y(vT) + \partial_z(wT) - \partial_z(v'_t \partial_z T) - \partial_x(A_h^T \partial_x T) - \partial_y(A_h^T \partial_y T) = \frac{\partial_z I}{c'_p \rho_0}, \quad (1)$$

kus  $\partial_t$  on osatuletis aja järgi,  $u$ ,  $v$  ja  $w$  vee liikumise kiirused vastavalt  $x$ ,  $y$  ja  $z$  suunas ning  $v'_t$  ja  $A_h^T$  vastavalt vertikaalne ja horisontaalne turbulentne viskoossus. Avaldise 1 paremal pool on päikesekiirguse neeldumine, kus  $I$  on kiirgus sügavusel  $z$ ,  $c'_p$  on vee erisoojusmahtuvus ning  $\rho_0$  vee tihedus normaalarõhul.

Samuti on vaja ääritingimusi ehk avaldisi pinna jaoks. Temperatuuri avaldis pinna jaoks on

$$v'_t \partial_z T = \frac{Q_s + Q_l + Q_b}{c'_p \rho_0}, \quad (2)$$

kus  $Q_s$ ,  $Q_l$  ja  $Q_b$  märgivad soojusvoogu tingituna vastavalt soojusvahetusest atmosfääriga, neeldumisest ja (pikalainelisest) tagasikiirgumisest.

## 1.2 ERGOM

ERGOM (*Ecological Regional Ocean Model*) on Läänemere jaoks loodud biogeokeemiline mudel, mis hõlmab endas lämmastiku- ja fosforiringet.

Joonisel 1 on toodud toitainete ringlus ERGOMi mudelis. Fütoplankton ehk taimhõljum on jagatud kolme gruppi: tsüanobakterid ehk sinivetikad, diatomeed ehk ränivetikad ning flagellandid ehk viburloomad. Nende olulisus seisneb algtoodangus (ingl k *primary production*), mis tähendab orgaanilise aine valmistamist, luues sellega toiduahela vundamendi. Kuna algtoodangu tootmine toimub peamiselt fotosünteesi kaudu, siis on suur roll valgusel. Sinivetikad töötlevad

atmosfäärist vette lahustunud lämmastiku ( $N_2$ ) lämmastikusidumise (ingl k *nitrogen fixation*) protsessi käigus ammooniumiks ( $NH_4^+$ ), mida on ülejäänud fütoplanktonil lihtsam tarbida. Seetõttu on sinivetikas justkui üks süsteemi lämmastikuallikatest. Samuti tarbib fütoplankton nitraati ( $NO_3^-$ ), mis tekib nitrifikatsiooni –ammooniumi nitraadiks muutumise –käigus. Loomhõljum toitub fütoplanktonist, mille abil liigub fütoplanktoni lämmastik zooplanktoni koosseisu. Nii loomse kui taimse kõljumi suremisel tekkiva laguprodukti ehk detriidi mineraliseerumise tulemusena tekib ammoonium, fosfaat ( $PO_4^{3-}$ ) ja lämmastik ning hapniku juuresolekul toimub taas nitrifikatsioon. Osa tekkinud nitraadist töödeldakse denitrifikatsiooni käigus uuesti lämmastikuks ning ringlus algab otsast peale. [3, 4]

Järgnev kirjeldus mudeli protsesside kohta on võetud artiklist [5]. Nitraadi ajaline muutus on kirja pandud kui

$$\frac{dN}{dt} = -\text{tarbimine} + \text{nitrifikatsioon} - \text{denitrifikatsioon} + \text{välised allikad}, \quad (3)$$

kus väliste allikate all on silmas peetud jõgede sissevoolust ja atmosfäärist tulevat lämmastikku. Nitrifikatsioon kiirus  $nf$  sõltub hapniku kontsentratsioonist  $O$  ja vee temperatuurist  $T$ :

$$\text{nitrifikatsioon} = nf \times A = \frac{O}{0.01 + O} \times 0.1 \exp(0.11T), \quad (4)$$

kus  $A$  on ammooniumi kontsentratsioon.

Denitrifikatsioon leiab aset väga madala hapniku kontsentratsiooni korral ning avaldub valemiga

$$\text{denitrifikatsioon} = f(N) \times \left( r_D D + \frac{r_S S}{H_{\text{pohi}}} \right), \quad (5)$$

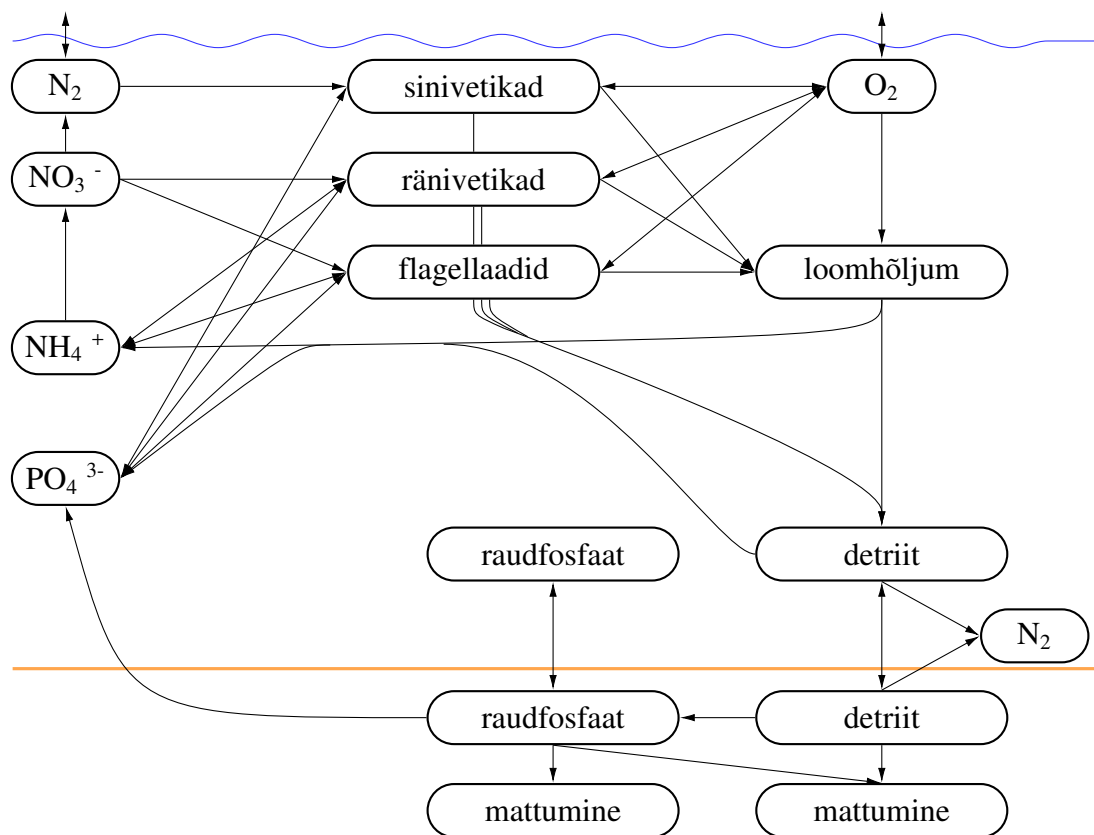
kus  $f(N) = 5.3 \times N^2 / (0.001 + N^2)$ , detriidi  $D$  ümbertöötlemise kiirus  $r_D = 0.003 \exp(0.15T)$ , setete  $S$  ümbertöötlemise kiirus  $r_S = 0.002 \exp(0.175T)$  ning põhjalähedase kihi paksus  $H_{\text{pohi}}$ . Fütoplanktoni gruppide sinivetikate, ränivetikate ja flagellaatide jaoks kehtivad vastavalt võrrandid

$$\frac{dF_1}{dt} = r_1 F_1 - h_1 F_1 - d_1 F_1 - G_1 \frac{F_1}{F_{\text{sum}}} Z - w_1^{\text{upp}} \frac{\partial F_1}{\partial z}, \quad (6)$$

$$\frac{dF_2}{dt} = r_2 F_2 - h_2 F_2 - d_2 F_2 - G_2 \frac{F_2}{F_{\text{sum}}} Z - w_2^{\text{upp}} \frac{\partial F_2}{\partial z}, \quad (7)$$

$$\frac{dF_3}{dt} = r_3 F_3 - h_3 F_3 - d_3 F_3 - G_3 \frac{F_3}{F_{\text{sum}}} Z, \quad (8)$$

kus  $r_i$  on kasvu kiirus,  $F_i$  fütoplanktoni grupp,  $h_i$  hingamise kiirus,  $d_i$  suremus,  $G_i$  tarbimise tempo (loomhõljumi  $Z$  toitumine) ning  $w_i^{\text{upp}}$  uppumise kiirus. Sisuliselt on iga grupi ajaline muutus võrdne nende kasvukiirusega, millest lahutatakse maha hingamine, suremus nii loomulikult teel kui zooplanktoni tõttu ning vetikate puhul ka nende uppumine. Fütoplanktoni ning mudeli teiste võrrandite täpsem kirjeldus leidub artiklis [3]



**Joonis 1.** Muutujad ja protsessid ERGOMi mudelis. Sinine laineline joon märgib merepinda, oranž sirgjoon merepõhja. Joonis põhineb ERGOMi kodulehel toodud joonisel, mida on kohandatud vastavalt simulatsiooni seadistustele.

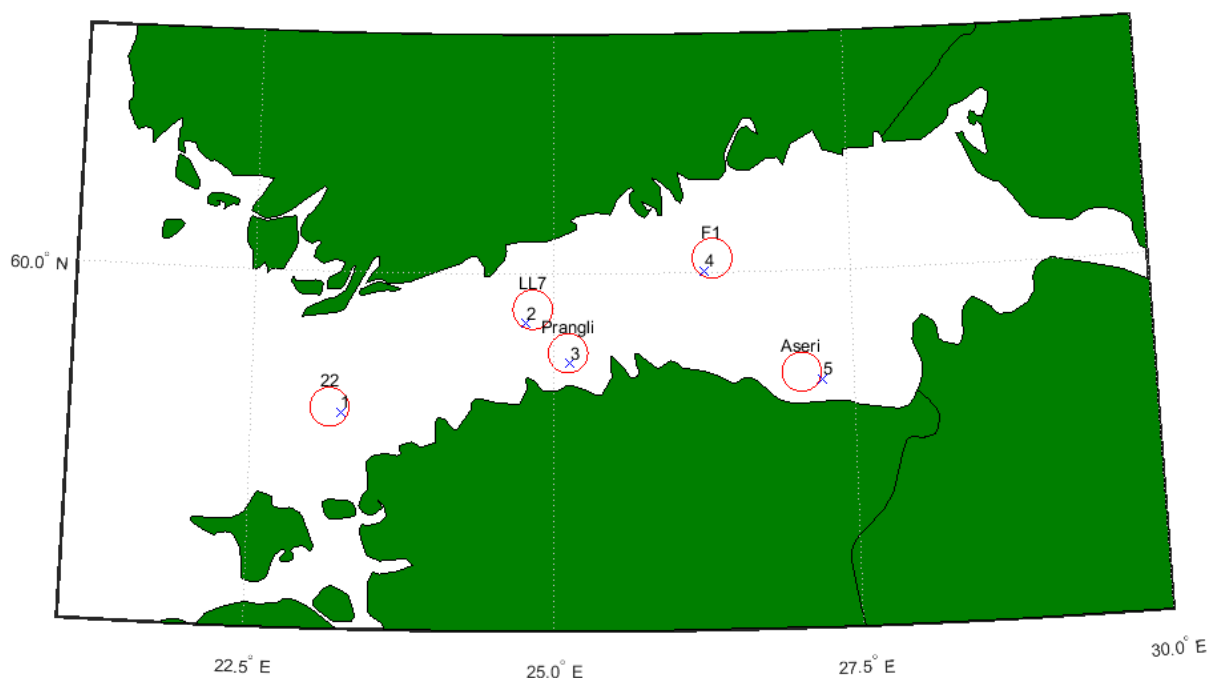
Käesolevas töös kasutatavad mudeli andmed saadi simulatsioonist, kus kasutati ühildatud GETM-ERGOM mudelit. Mudeli füüsikalise osa (GETM) seadistused on toodud artiklis [6] ning bioloogilise osa (ERGOM) seadistused artiklis [5].

## 2 Metoodika

### 2.1 Andmed

#### Jaamad

Mudelit valideeriti viie jaama andmete põhjal. Mudeli täpsusele objektiivse hinnangu saamiseks valiti jaamad, mille kohta leidsu piisavalt mõõtmisandmeid ning mis asusid erinevates Soome lahe piirkondades.



**Joonis 2.** Valitud jaamad. Sinine rist märgib mudeli jaama asukohta ning punane ring piirkonda, mille sees asuvad vastavad vaatlusandmed.

Joonisel 2 on näidatud jaamade asukohad kaardil. Vaatlusjaam 22 asub Soome lahe suudmes, 20 km Osmussaarest loodes ja 45 km Soomest lõunas. LL7 asub Tallinna ja Helsinki vahel, Tallinnast 43 km ning Helsinkist 36 km kaugusel. Prangli vaatlusjaam asub Prangli saarest 7 km kirdes, 22 km Juminda poolsaarest lääneloodes. F1 asub keset Soome lahte, Pärissaare poolsaarest 57 km kirdes ja Suursaarest 35 km läänes. Aseri vaatlusjaam on kõige rannikulähedasem, asudes Aserist 18 km kirdes.

Mudeli jaamad valiti selliselt, et sealne merepõhja pinnavorm esindaks vaatlusjaamaga tabada püütud pinnavormi või sügavusvahemikku.

Jaamade täpsemad koordinaadid ning sügavus mudeli jaamas on toodud tabelis 1.

**Tabel 1.** Valitud jaamad, nende koordinaadid ja sügavused. Vaatlusjaamade raadiused jäid vahemikku 9,118 km kuni 9,285 km.

| Vaatlusjaamad |                    | Mudeli jaamad |                    |             |
|---------------|--------------------|---------------|--------------------|-------------|
| Nimi          | Koordinaadid       | Nimi          | Koordinaadid       | Sügavus (m) |
| 22            | 59,4300N, 23,1500E | 1             | 59,4088N, 23,2451E | 95.3        |
| LL7           | 59,8500N, 24,8300E | 2             | 59,7916N, 24,7761E | 65.0        |
| Prangli       | 59,6672N, 25,1246E | 3             | 59,6252N, 25,1422E | 70.4        |
| F1            | 60,0600N, 26,3400E | 4             | 60,0080N, 26,2738E | 60.8        |
| Aseri         | 59,5742N, 27,0650E | 5             | 59,5419N, 27,2390E | 37.6        |

## Mõõtmisandmed

Mõõtmisandmed koguti rakenduse Nest abil. Nest on Stockholmi Ülikooli Baltic Nest Institute'i otsustetegemise abisüsteem (ingl k *decision support system*), mis sisaldab Läänemerd hõlmavaid mudeleid ja mõõtmiste andmebaasi [7]. Andmete saamiseks valiti rakenduses soovitud jaam (vt joonis 3) ja määrati ajavahemik (aastad 1966 kuni 2005). Kuude keskmiste väärtuste saamiseks valiti keskmistamise intervalliks 30 päeva. Valikute kinnitamisel väljastas programm mõõtmisandmed, mille sai arvutisse laadida CSV-failidena, ja info jaama kohta.

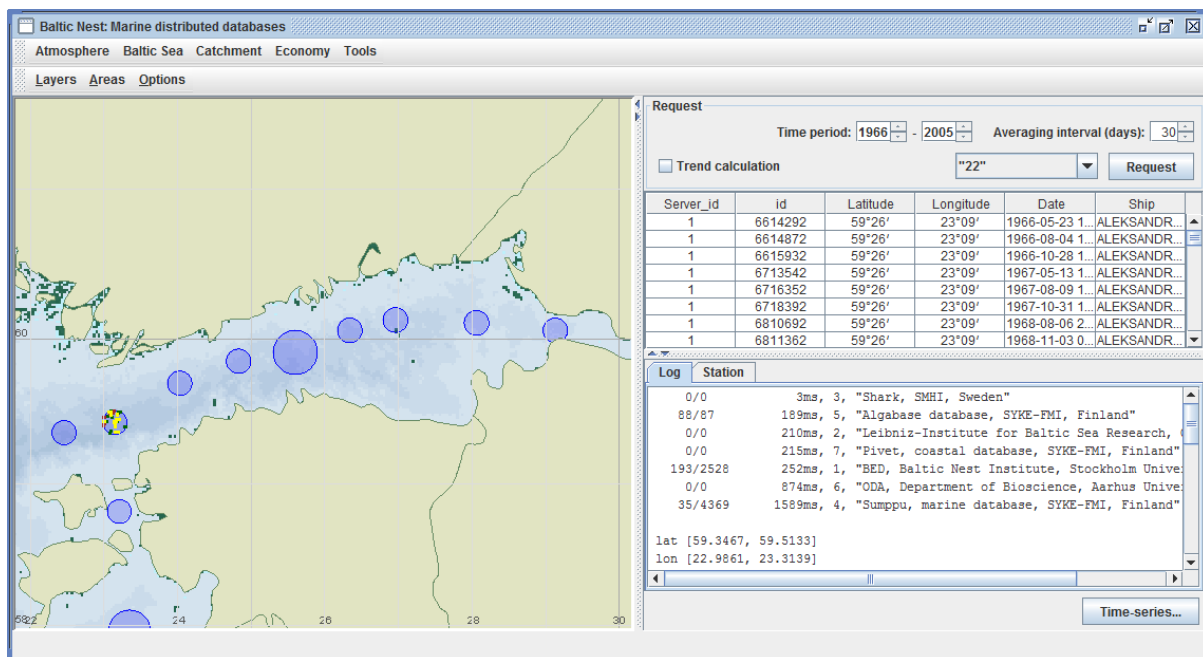
Erinevalt jaamadest 22, LL7 ning F1 ei olnud süsteemis vaikimisi Prangli ega Aseri jaama. Viimaste jaoks kasutati rakenduse võimalust uue mõõtmiste ala defineerimiseks. Uute jaamade suurus valiti võimalikult lähedale olemasoleva kolme jaama suurusele.

Väljastatud mõõtmisandmetest kasutati järgmisi füüsikalisi ja biogeokeemilisi välju:

- temperatuur (°C),
- nitraadis ( $\text{NO}_3$ ) ja nitritis ( $\text{NO}_2$ ) sisalduv lämmastik ( $\mu\text{mol}_\text{N}/\text{l}$ ),
- klorofüll ( $\mu\text{g}/\text{l}$ ).

Nestist saadi iga jaama iga välja kohta CSV-fail, mis teisendati Excelis tabelkujule. Tabeli esimeses veerus olid sügavused meetrites (1.2, 5, 10, 15, 20.8, 29.5, 40, 50, 60, 70, 80, 90, 100, 121.1) ning esimeses reas kuud 1966. aasta jaanuarist 2005. aasta detsembrini, kokku 480 kuud. Tabelid salvestati Matlabi maatriksiteks, kus eraldati teineteisest sügavuste veerg, kuupäevade rida ning mõõtmisandmed. Suvine suur UV-kiirgus vähendab klorofüllü väärtust pinnal ega anna





**Joonis 3.** Pilt rakendusest Nest. Sinised ringid märgivad Läänemere peamisi vaatlusjaamu.

valideerimiseks usaldusväärseid väärtusi, kuna modelis ei ole UV-kiirguse mõju fütoplanktonile arvesse võetud. Seetõttu, kuna Soome lahe pinnakiht on 5 m ulatuses peaaegu alati segunenud [8], väljendavad antud töös pinnakihti 5 m sügavusel mõõdetud andmed.

Kõigil Nesti väljastatud tabelitel on sama formaat ehk neis kõigis on read eespool loetletud sügavuste kohta, kuid kõik viis töös käsitletud jaama olid madalamad kui tabeli maksimaalne, 121.1 m. Seetõttu eemaldati maatriksitest lisaks pinna andmetele ka read, millele vastav sügavus oli suurem kui mudeli jaama sügavus.

Lämmastikisisalduse teisendamiseks ühikult  $\mu\text{mol/l}$  ühikule  $\text{mmol/m}^3$  ei olnud tarvis mõõtmisandmeid muuta, kuna need on võrdväärsed.

Eraldi tuli läheneda klorofüllü andmestikule, kuna sellele ei ole otseselt vastavaid mudeli andmeid. Klorofüll on biomolekul, millel on suur roll fotosünteesis: see on oluline, et taim saaks valgusenergiat neelata. Kuna fütoplanktoni lämmastikutarbimine sõltub valgustingimustest, on fütoplanktoni koosseisu kuuluv lämmastik seotud klorofüllü kontsentratsiooniga fütoplanktonis. Mõõtmisandmetes on klorofüllü kontsentratsioon fütoplanktonis, kuid selle mõõtmise tegelik eesmärk on tavaliselt saada teada fütoplanktonis sisalduva orgaanilise süsiniku kogus [9]. Selle leidmiseks korrutatakse klorofüllü kontsentratsiooni süsiniku ja klorofüllü suhtega C:Chl.

Redfieldi suhte kohaselt on (füto)planktonis sisalduva orgaanilise süsiniku, lämmastiku ja fosfori osakaalud 106:16:1 (C:N:P) [10]. Seega oleks lihtne viia omavahel vastavusse mõõdetud süsiniku kontsentratsioon ning mudeli ennustatud lämmastiku kontsentratsioon fütoplanktonis (summa sinivetikates, ränivetikates ja flagellaatides sisalduvast lämmastikust), korrutades mõõtmisandmeid 16/106-ga.

Keerukust lisab aga, et C:Chl ise ei ole konstantne suurus. See sõltub fütoplanktoni kasvu kiirusest, valgustingimustest ning liikide koosseisust, mis võivad omakorda sõltuda toitainete kogustest. Samadel põhjustel ei ole C:Chl konstantne ka ühe veesamba piires. Läänemere põhjaosa kohta tehtud uuringus leiti, et C:Chl varieerus vahemikus 5 kuni 70, olles madalaim sügistalvisel perioodil ning kõrgeim kevadel. [11]

Lisaks nimetatud mõjuritele on suhe erinev ka fütoplanktoni eri gruppide jaoks, olles näiteks flagellaatides suurem kui ränivetikates. Kõiki tegureid arvestades võib C:Chl olla nii 10 kui ka 250 või rohkem. [9]

Käesolev töö ei keskendunud klorofüllis ja süsiniku suhte leidmisele. Seetõttu leiti, mis oleks lämmastikusisaldus klorofüllis, kui suhe C:Chl oleks 1. Selleks korrutati mõõtmisandmed Redfieldi suhtest tulenevalt 0.151-ga ning teisendati ühikust  $\mu\text{g/l}$  ühikusse  $\text{mmol/m}^3$ , kasutades lämmastiku molaarmassi 14.007 g/mol:

$$\frac{\mu\text{g}}{1} = \frac{\mu\text{g}}{14.007 \frac{\text{g}}{\text{mol}}} \times \frac{1}{1} = \frac{\mu\text{mol}}{14.0071} = \frac{\text{mmol}}{14.007 \text{ m}^3}. \quad (9)$$

Teisendusega saadud väärtused on erinevad tegelikkusest C:Chl korda, kuid on mudeli andmete suurusjärgule lähemal.

Pärast nimetatud tegevusi salvestati kõik maatriksid edasiste arvutuste jaoks.

## Mudeli andmed

Päevade keskmistest mudeli andmetest võeti mudeli jaamade kohta välja järgmised parameetrid:

- ajavektor sammuga üks päev alates 01.01.1966 kuni 31.12.2005 (kokku 14610 päeva),
- jaama pikkus- ja laiuskraad,
- jaama sügavus (m),
- sügavusvektor sammuga 5 m,

- temperatuur ( $^{\circ}\text{C}$ ),
- nitraadis ( $\text{NO}_3$ ) ja nitritis ( $\text{NO}_2$ ) sisalduv lämmastik ( $\text{mmol}_\text{N}/\text{m}^3$ ),
- ränivetikates sisalduv lämmastik ( $\text{mmol}_\text{N}/\text{m}^3$ ),
- flagellaatides sisalduv lämmastik ( $\text{mmol}_\text{N}/\text{m}^3$ ),
- sinivetikates sisalduv lämmastik ( $\text{mmol}_\text{N}/\text{m}^3$ ).

Maatriksites vastas iga rida ühe päeva andmetele ning iga veerg ühe sügavuse andmetele põhjast pinnani.

Mudeli andmetest ekstraheeriti mõõtmisandmetele vastavad sügavused ning seejärel leiti igas jaamas igal sügavusel kuude keskmised vastavalt kuu numbrile. Mudeli andmete maatriksid transponeeriti, et sügavuse ja aja dimensioonid oleksid vastavuses vaatlusandmete dimensioonidega. Kõik maatriksid salvestati MAT-faili.

## 2.2 Valideerimismeetodid

Mudeli valideerimisel kasutati nii statistilisi näitajaid kui ka graafilist võrdlust. Kuna statistikud hindavad mudeli sooritust erinevalt [12], võeti usaldusväärse hinnangu andmiseks kõrvuti kasutusele mitu statistikut. Eelistati ühikuta suurusi nagu korrelatsioonikordaja, Nash-Sutcliffe'i efektiivsus, suhteline viga ja kulufunktsioon, kuna nende väärtustele saab vastavusse seada skaala (vt tabel 2), mille alusel mudeli sooritust subjektiivselt hinnata. Kuna statistikute väärtusi võivad üksikud suured kõrvalekalded palju mõjutada (eriti väikese arvu mõõtmiste korral), siis kasutati paralleelselt ka graafilist võrdlust, kus keskenduti regressioonsirgega hajuvusdiagrammidele ning aegridadele.

### 2.2.1 Statistilised näitajad

#### Lineaarne korrelatsioonikordaja

Korrelatsioonikordaja on ühikuta suurus lõigus  $[-1, 1]$ , mis näitab, kui tugev on lineaarne sõltuvus kahe suuruse, käesolevas töös mõõtmis- ja mudeli andmete vahel. Parimal juhul, kui mudel on täpne ja andmed kattuvad, on kordaja väärtus 1, negatiivne väärtus viitab mudeli ja reaalse maailma andmete pöörd sõltuvusele. Nullilähedase korrelatsiooni korral on kaks suurst

omavahel lineaarselt sõltumatud. Korrelatsioonikordaja  $r$  avaldub valemiga

$$r = \frac{\sum_{i=1}^n (V_i - \bar{V})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (V_i - \bar{V})^2 \sum_{i=1}^n (M_i - \bar{M})^2}}, \quad (10)$$

kus siin ja edaspidi

$n$  – mõõtmiste arv,

$V, \bar{V}$  – vastavalt vaatlusandmed ja vaatlusandmete keskmine,

$M, \bar{M}$  – vastavalt mudeli andmed ja mudeli andmete keskmine. [12, 13, 14]

### Nash-Sutcliffe'i efektiivsus

Nash-Sutcliffe'i efektiivsus määrab, kui hästi kattub vaatlusandmete *versus* mudeli andmete graafik  $y = x$  joonega. Selle väärtus on poollõigis  $(-\infty, 1]$ , kus heaks tulemuseks peetakse positiivseid väärtusi. Null või negatiivne väärtus tähendab, et vaatlusandmete keskmine ennustab reaalsel maailma paremini kui mudel. Nash-Sutcliffe'i efektiivsust  $E$  arvutatakse valemiga

$$E = 1 - \frac{\sum_{i=1}^n (M_i - V_i)^2}{\sum_{i=1}^n (V_i - \bar{V})^2} = 1 - \frac{RMS E^2}{\sigma_v^2}, \quad (11)$$

kus  $RMS E$  on mudeli ja mõõtmisandmete vaheline standardhälve ning  $\sigma_v$  on mõõtmisandmete standardhälve. [15]

### Suhteline viga

Mudeli ja mõõtmisandmete vahe jagamisel mõõtmisandmetega saab suhtelise vea, mis näitab, kui palju mudel keskmiselt üle- või alahindab reaalse maailma väärtusi. Positiivne väärtus tähendab, et mudel alahindab, ja negatiivne, et ülehindab. Täpsete ennustuste korral suhteline viga  $\delta = 0$ . Käesolevas töös leitakse suhteline viga protsentides ning  $\delta$  arvutatakse valemiga [15]

$$\delta = \frac{\sum_{i=1}^n (V_i - M_i) \times 100}{\sum_{i=1}^n V_i}. \quad (12)$$

### Kulufunktsioon

Kulufunktsiooni (ingl k *cost function*) arvutatakse valemiga

$$K = \left| \frac{\bar{M} - \bar{V}}{\sigma_v} \right| \quad (13)$$

ning see näitab, mitme mõõtmiste standardhälbe kaugusele jääb mudeli keskmine mõõtmisandmete keskmisest.  $K$  väärtuste tõlgendamiseks kasutatakse skaalat tabelis 2, kus „väga hea” tähendab, et mudeli keskmine ei erine vaatlusandmete keskmisest rohkem kui ühe (vaatlusandmete) standardhälbe võrra. [16]

**Tabel 2.** Mudeli osavuse hindamine erinevate mõõdikute järgi.  $r$  jaoks võeti kasutusele skaala, mida kasutati artiklis [17] veelahkme mudeli veevoolu valideerimiseks.

| Hinnang       | $r$ [17]             | $E$ [18]             | $\delta$ [%] [19, 20]   | $K$ [21]       |
|---------------|----------------------|----------------------|-------------------------|----------------|
| Väga hea      | $0.92 < r \leq 1.00$ | $0.65 < E \leq 1.00$ | $ \delta  < 10$         | $0 \leq K < 1$ |
| Hea           | $0.87 < r \leq 0.92$ | $0.50 < E \leq 0.65$ | $10 \leq  \delta  < 20$ | $1 \leq K < 2$ |
| Rahuldav      | $0.80 < r \leq 0.87$ | $0.20 < E \leq 0.50$ | $20 \leq  \delta  < 40$ | $2 \leq K < 3$ |
| Mitterahuldav | $r < 0.80$           | $E \leq 0.20$        | $ \delta  \geq 40$      | $K \geq 3$     |

### 2.2.2 Graafiline võrdlus

#### Hajuvusdiagramm ja regressioonsirge

Hajuvusdiagramm on graafik, mille ühele teljele kantakse mõõtmisandmed ning teisele mudeli andmed, käesolevas töös vastavalt horisontaal- ja vertikaalteljele. Lisaks leiti graafiku punktidele regressioonsirge. Regressioonsirge tõus hindab mudeli ja mõõtmisandmete suhet, mis parimal juhul on 1 [15]. Algordinaadi väärtus on regressioonsirge põhjal keskmine väärtus, mida mudel ennustab, kui mõõdetud väärtus on 0. Regressioonsirge juures võetakse toeks ka determinatsioonikoefitsient  $D = r^2$ , mis näitab, kui palju hajuvusest on seletatav korrelatsiooniga [14], s.t kui hea lähendusega on leitud regressioonsirge.

**Tabel 3.** Regressioonsirge lähenduse hinnang determinatsioonikoefitsendi järgi.

| Hinnang       | $r^2$ [17]           |
|---------------|----------------------|
| Väga hea      | $0.85 < r \leq 1.00$ |
| Hea           | $0.75 < r \leq 0.85$ |
| Rahuldav      | $0.65 < r \leq 0.75$ |
| Mitterahuldav | $r < 0.65$           |

#### Aegread

Mõõtmis- ja mudeli andmete aegridade kuvamine samale graafikule annab hea ülevaate nende omavahelisest ajalisest kokkulangevusest ja väärtuste ulatusest. Aegread annavad lisainformatsiooni eriti seal, kus statistikute järgi on mudeli sooritus kehv, kuna aegridadelt tulevad välja üksikud suured kõrvalekalded ning mudeli ja mõõtmisandmete vaheline ajaline nihe, mis võivad

statistilisi näitajaid palju mõjutada.

## 2.3 Arvutused

Arvutused teostati Matlabis eelnevalt salvestatud maatriksitega. Iga välja juures leiti enne arvutuste tegemist mõõtmisandmete maatriksis NaN-id (ingl k *not a number*) ehk puuduvad andmed ning eemaldati vastavate indeksitega väärtused nii mõõtmiste kui ka mudeli maatriksist. Igas jaamas leiti statistikud temperatuurile, lämmastikussisaldusele ning klorofüllü kontsentratsioonile, kusjuures mõõtmiste sügavused 20.8 m ja 29.5 m viidi vastavusse mudeli sügavustega 20 m ja 30 m. Iga välja kohta leiti arvutuses kasutatud andmepaaride arv ( $n$ ), mõõtmisandmete keskmine ( $\bar{V}$ ) ja standardhälve, mudeli andmete keskmine ( $\bar{M}$ ) ja standardhälve, mõõtmisandmete väärtuste vahemik, andmepaaride ruutkeskmine hälve ( $RMS E$ ), korrelatsioonikordaja ( $r$ ), Nash-Sutcliffe'i efektiivsus ( $E$ ), suhteline viga ( $\delta$ ) ning kulufunktsioon ( $K$ ). Lisaks leiti loetletud suurused kokkuvõtvalt ka kõikide jaamade andmetele. Klorofüllü jaoks leiti igas jaamas keskmine süsiniku ja klorofüllü suhe.

Mõõtmis- ja mudeli andmete keskmise ning standardhälbe leidmisel kasutati Matlabi vastavaid funktsioone `mean` ja `std`, mõõtmisandmete väärtuste vahemiku leidmiseks funktsioone `min` ja `max` ning korrelatsiooni leidmiseks funktsiooni `[r,p] = corrcoef(V,M)`, kus

$r$  – korrelatsioonikordaja,

$p$  – p-väärtus.

P-väärtus on lõigus  $[0, 1]$  ning näitab, kui suur on tõenäosus teha viga, öeldes, et kahe suuruse vahel on korrelatsioon. P-väärtus aitab otsustada, kas on mõistlik võtta vastu otsus, et suuruste vahel on korrelatsioon. Selleks kasutatakse olulisuse nivood  $\alpha$ : kui  $p < \alpha$ , siis võib väita, et eksisteerib korrelatsioon.  $p \geq \alpha$  korral ei või väita, et kahe suuruse vahel korrelatsioon puudub. Käesolevas töös kasutatakse nivood  $\alpha = 0.05$ . [14]

Ruutkeskmine hälve ehk ruuthälve  $RMS E$  (ingl k *root mean square error*) leiti valemiga [12]

$$RMS E = \sqrt{\frac{\sum_{i=1}^n (M_i - V_i)^2}{n}} \quad (14)$$

ning selle puhul kehtib, et 68.27% mudeli ennustustest  $M_i$  mahub vahemikku  $V_i \pm RMS E$ , 95.45% vahemikku  $V_i \pm 2RMS E$  ning 99.73% vahemikku  $V_i \pm 3RMS E$  [14].

Ülejäänud statistikute jaoks defineeriti eelnevalt funktsioonid vastavalt valemitele (11)-(13).

### 3 Tulemused ja arutelu

Jutumärkides antud hinnangud põhinevad tabelitel 2 ja 3.

#### 3.1 Temperatuur

Tabelis 4 on toodud temperatuuri kohta arvutatud statistikud igas jaamas eraldi ning kõikide jaamade andmete peale kokku. Võrreldavate andmepaaride hulk varieerus 158st 2455ni, kokku 6533 andmepaari.

**Tabel 4.** Statistilised näitajad temperatuuri kohta kõikides jaamades.  $n$  on andmepaaride arv,  $\overline{T_m}$  ja  $\overline{T_v}$  on vastavalt mudeli andmete ja mõõtmisandmete temperatuuri keskmine koos standardhälbega.

| Jaam    | $n$  | $\overline{T_m}$ [°C] | $\overline{T_v}$ [°C] | $T_v$ vahemik [°C] | $RMSE$ [°C] | $r$  | $p$    | $E$  | $\delta$ [%] | $K$  |
|---------|------|-----------------------|-----------------------|--------------------|-------------|------|--------|------|--------------|------|
| 22      | 1541 | 6.75±4.62             | 5.69±4.15             | −0.30...20.50      | 2.22        | 0.91 | < 0.05 | 0.71 | −18.70       | 0.26 |
| LL7     | 2455 | 6.69±4.77             | 5.34±4.41             | −0.54...20.45      | 2.37        | 0.91 | < 0.05 | 0.71 | −25.22       | 0.31 |
| Prangli | 1174 | 6.70±4.56             | 5.48±4.45             | −0.13...20.89      | 2.34        | 0.90 | < 0.05 | 0.72 | −22.41       | 0.28 |
| F1      | 1205 | 6.19±5.02             | 5.10±4.69             | −0.59...22.24      | 2.26        | 0.92 | < 0.05 | 0.77 | −21.42       | 0.23 |
| Aseri   | 158  | 6.66±6.37             | 5.92±6.00             | −0.21...18.58      | 2.03        | 0.95 | < 0.05 | 0.88 | −12.57       | 0.12 |
| Kokku   | 6533 | 6.61±4.79             | 5.42±4.46             | −0.59...22.24      | 2.30        | 0.91 | < 0.05 | 0.73 | −22.10       | 0.27 |

Mudeli ja mõõtmisandmete keskmised ning suhteline viga  $\delta$  viitavad mudeli ülehindamisele. Suhtelise vea järgi on mudeli ennustusvõime jaamades LL7, Prangli ja F1 „rahuldav” (−21.42...−25.22%) ning jaamades 22 ja Aseri „hea” (vastavalt −18.70% ja −12.57%). Kõikide jaamade arvestuses hindab  $\delta$  mudelit „rahuldavaks”, kusjuures suhteline viga on −22.1% ehk mudel ennustab tegelikkusest keskmiselt 22.1% kõrgemat temperatuuri. Joonistelt 5 ning 6 on näha, et mudel ennustab põhjas temperatuuri kehvemini kui pinnal, mis võib olla suure suhtelise vea põhjuseks.

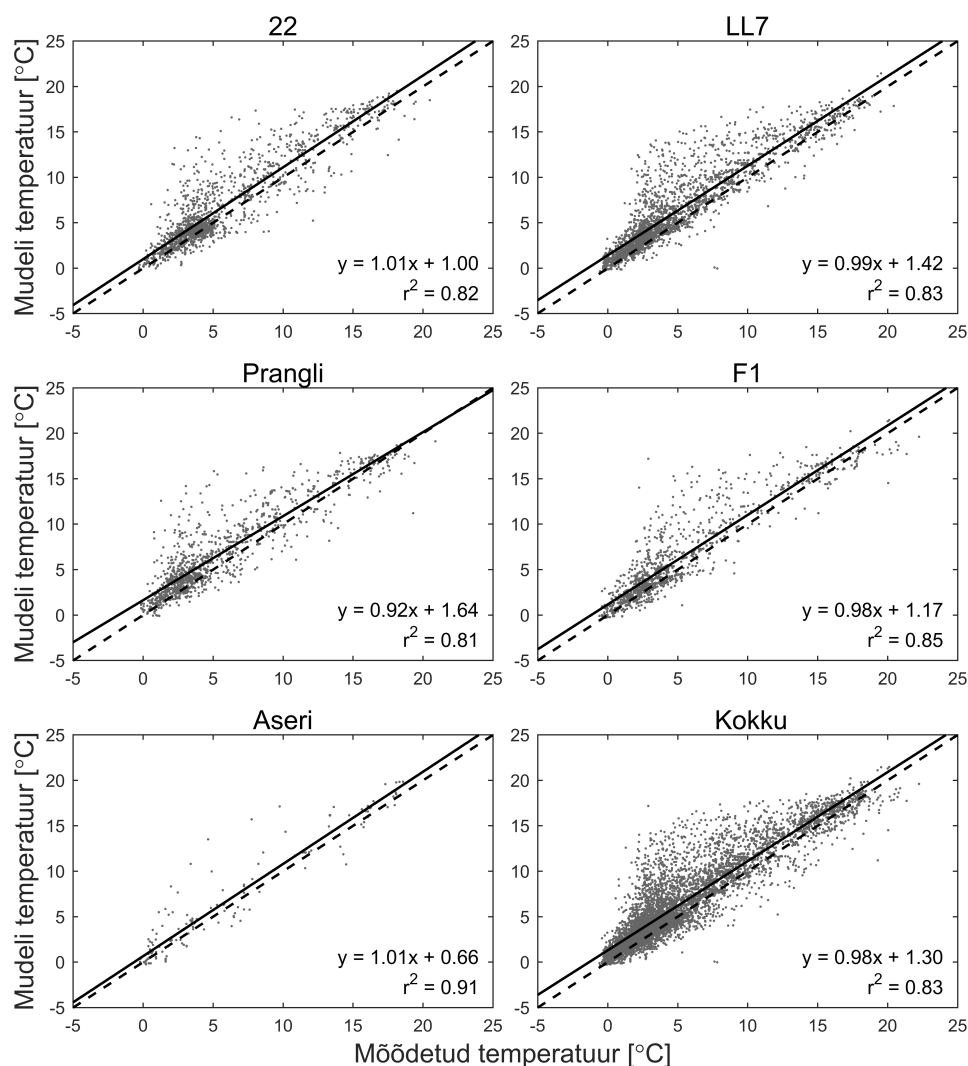
Korrelatsioonikordaja põhjal on mudeli sooritus jaamades 22, LL7 ja Prangli „hea” (0.90...0.91) ning jaamades F1 ja Aseri „väga hea” (0.92 ja 0.95). Kõiki temperatuuriandmeid arvestades võib mudelit hinnata „heaks” ( $r = 0.91$ ). Joonistelt 5 ning 6 on näha, et nii pinnal kui ka põhjas varieerub mudeli joon koos vaatlusandmete punktidega, mis kinnitab tugevat korrelatsiooni mudeli ja mõõtmisandmete vahel.

Nii Nash-Sutcliffe'i efektiivsuse  $E$  kui ka kulufunktsiooni  $K$  väärtused viitavad mudeli „väga heale” sooritusele igas jaamas eraldi ( $E$  0.71...0.88,  $K$  0.21...0.31) kui ka kõikide jaamade peale



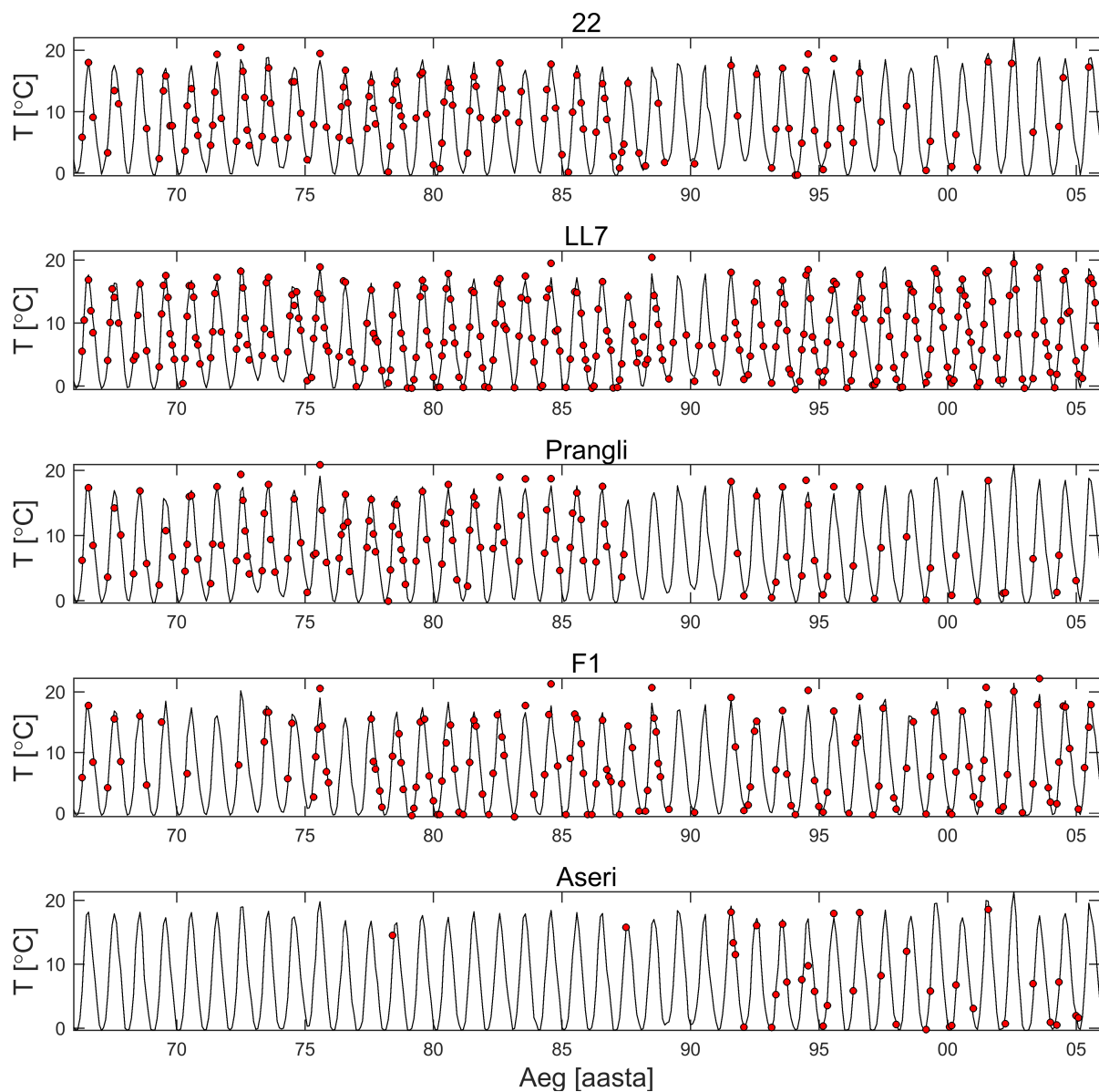
kokku ( $E = 0.73$ ,  $K = 0.27$ ).

Hajuvusdiagrammidelt joonisel 4 on näha, regressioonsirge on iga jaama korral  $y = x$  joonest ülalpool, mis viitab taas mudeli temperatuuri ülehindamisele, kuid sirge tõus on lähedal soovitud väärtusele 1, olles halvim ehk 0.92 Prangli jaamas ning mujal 0.98...1.01. Determinatsioonikoefitsendi põhjal on mudeli sooritus igas jaamas „väga hea”: korrelatsioon seletab 81...91% hajuvusest.

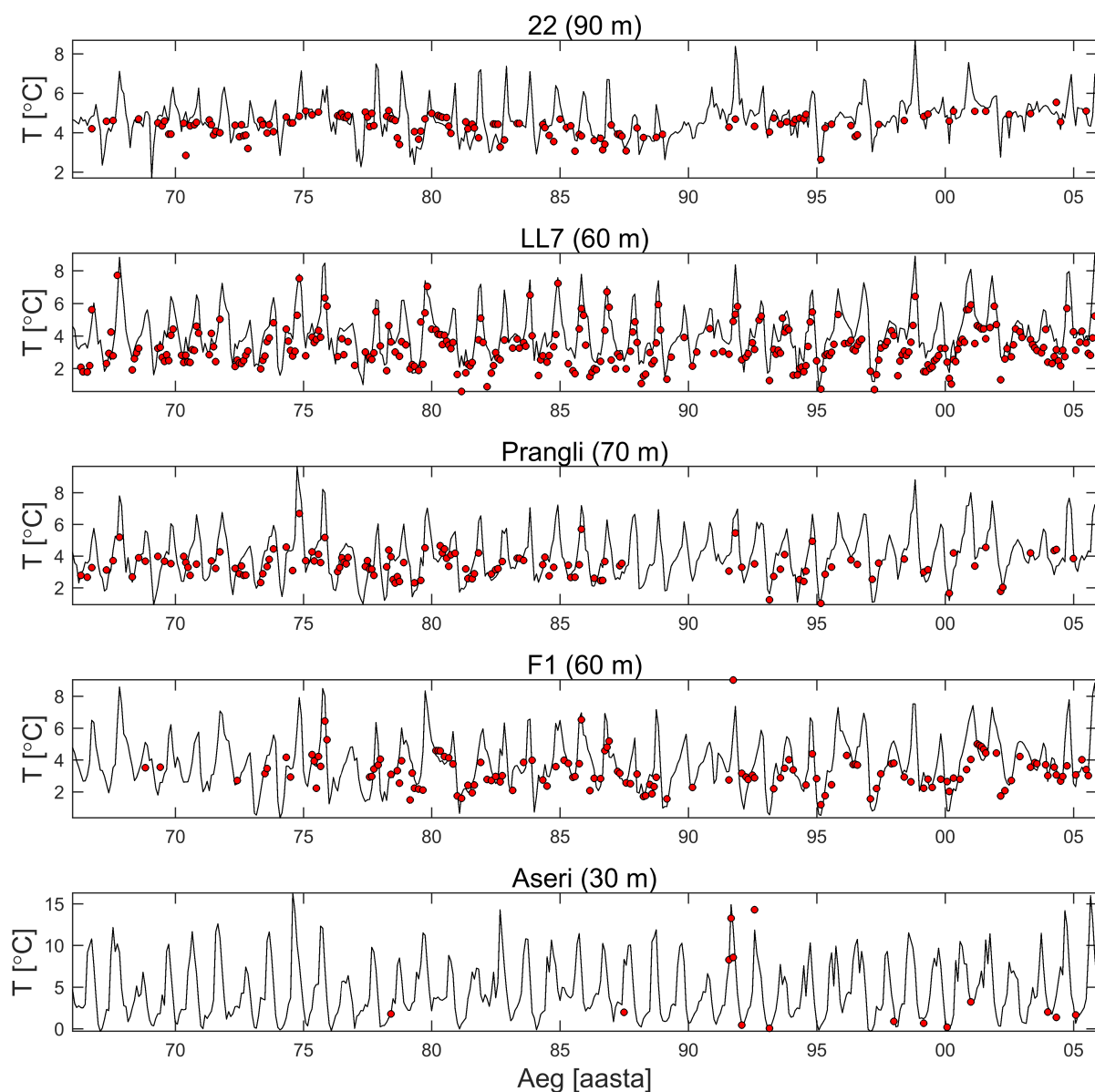


**Joonis 4.** Temperatuuri hajuvusdiagrammid. Katkendjoon märgib  $y = x$  joont, pidev joon regressioonsirget, mille võrrand on iga graafiku all paremas nurgas.

Nii statistikuid kui ka graafikuid arvesse võttes võib väita, et vaadeldud jaamades ennustab mudel temperatuuri väga hästi.



**Joonis 5.** Temperatuuri aegread pinnal. Joon märgib mudeli andmeid ning punktid mõõtmisandmeid.



**Joonis 6.** Temperatuuri aegread põhjas. Joon märgib mudeli andmeid ning punktid mõõtmisandmeid.

### 3.2 Nitraadis sisalduv lämmastik

Tabelis 5 on toodud lämmastikuisalduse kohta arvutatud statistikud igas jaamas ning kõikide jaamade andmete peale kokku. Võrreldavaid andmepaare oli 137...2037, viie jaama peale kokku 4774 paari.

Mudeli ja mõõtmisandmete keskmiste ning suhtelise vea põhjal saab öelda, et mudel üldiselt ülehindab lämmastikuisaldust merevees, ainsaks erandiks jaam F1. Vaid ühes jaamas (LL7) on mudeli osavus „väga hea” ( $\delta = -4.84\%$ ), jaamas F1 on mudel „hea” ( $\delta = 11.21\%$ ) ning

**Tabel 5.** Statistilised näitajad lämmastiku kohta kõikides jaamades.  $n$  on andmepaaride arv,  $\overline{N_m}$  ja  $\overline{N_v}$  on vastavalt mudeli andmete ja mõõtmisandmete lämmastikuisalduse keskmine koos standardhällbega.

| Jaam    | $n$  | $\overline{N_m}$ [mmol/m <sup>3</sup> ] | $\overline{N_v}$ [mmol/m <sup>3</sup> ] | $N_v$ vahemik [mmol/m <sup>3</sup> ] | $RMS E$ [mmol/m <sup>3</sup> ] | $r$  | $p$    | $E$   | $\delta$ [%] | $K$  |
|---------|------|---|---|--------------------------------------|--------------------------------|------|--------|-------|--------------|------|
| 22      | 844  | 4.64±3.60                               | 3.40±3.00                               | 0.00...16.38                         | 3.53                           | 0.51 | < 0.05 | -0.39 | -36.46       | 0.41 |
| LL7     | 2037 | 3.92±2.88                               | 3.74±3.05                               | 0.00...11.87                         | 2.92                           | 0.52 | < 0.05 | 0.08  | -4.84        | 0.06 |
| Prangli | 665  | 4.40±3.02                               | 3.56±3.05                               | 0.00...19.68                         | 3.26                           | 0.46 | < 0.05 | -0.14 | -23.48       | 0.27 |
| F1      | 1091 | 4.56±3.05                               | 5.14±3.73                               | 0.00...12.64                         | 3.07                           | 0.62 | < 0.05 | 0.32  | 11.21        | 0.15 |
| Aseri   | 137  | 5.28±3.78                               | 4.36±4.16                               | 0.00...13.28                         | 3.30                           | 0.68 | < 0.05 | 0.37  | -21.18       | 0.22 |
| Kokku   | 4774 | 4.30±3.13                               | 3.99±3.31                               | 0.00...19.68                         | 3.13                           | 0.53 | < 0.05 | 0.10  | -7.70        | 0.09 |

ülejäänud jaamades „rahuldav” ( $|\delta|$  -21.18...-36.46%). Kuigi jaamade lõikes varieeruvad hinnangud „rahuldavast” „väga heani”, siis kõikide jaamade andmete peale tuli suhteliseks veaks -7.70% ehk „väga hea”, mis on ilmselt tingitud väikesest suhtelisest veast jaamas LL7 ning positiivsest suhtelisest veast jaamas F1, mis suure mõõtmiste arvu tõttu tasakaalustavad kehva väärtust nagu  $\delta = -36.46\%$  (peaaegu „mitterahuldav”) jaamas 22.

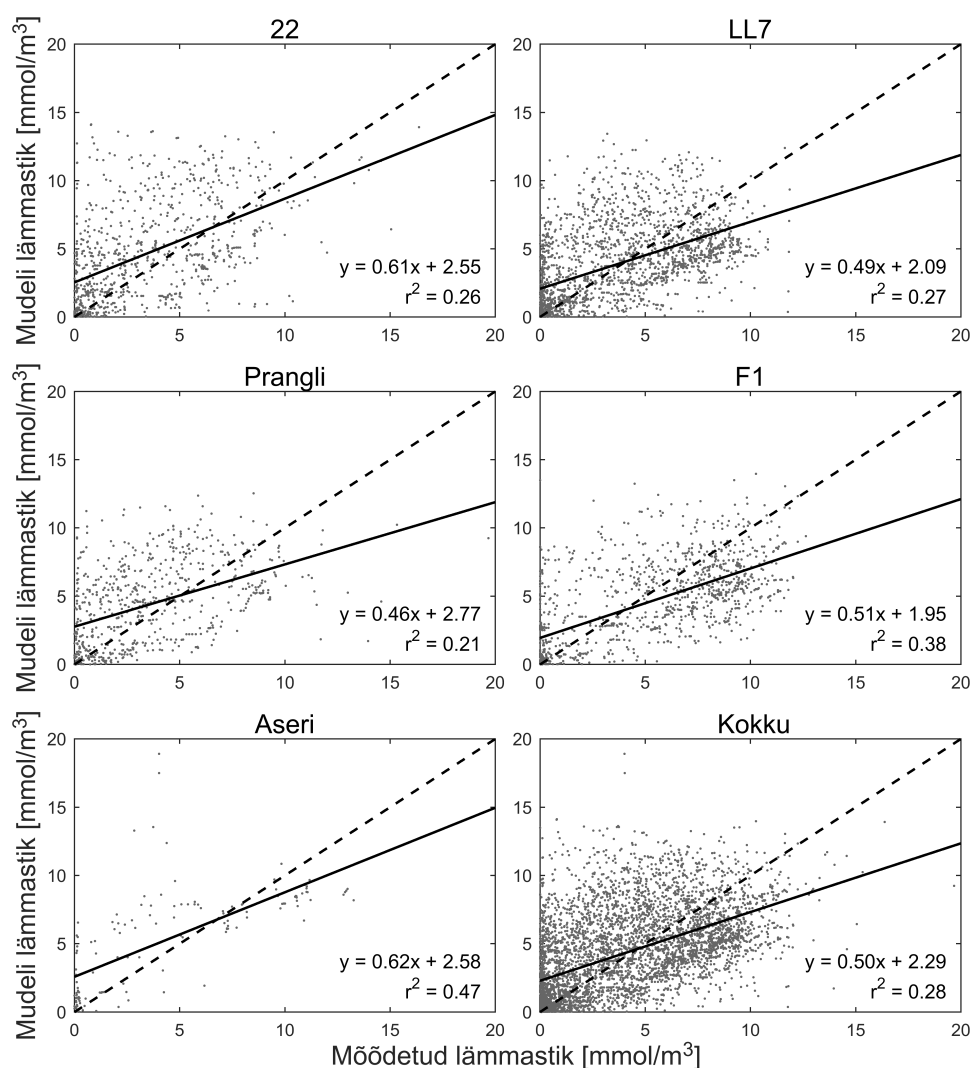
Joonistelt 8 ja 9 on näha, et pinnal kaldub mudel tegelikke väärtusi veidi alahindama (eriti talvist fotosünteesi mittetoimumisest tulenevat suuremat lämmastikuisaldust), kuid põhjas pigem ülehindama. Selline mudeli erinev käitumine pinna ja põhja ennustamisel võib samuti tingida keskmiselt hea tulemuse.

Joonisel 7 toodud lämmastiku hajuvusdiagrammidelt ei ole näha, et mudel üheski jaamas lämmastikku väga hästi ennustaks. On näha, et mudeli tulemused üldiselt küll kasvavad mõõtmisandmete kasvades, kuid punktid on väga hajusalt. Ka korrelatsioonikoefitsendi, mis on vahemikus 0.51...0.68, põhjal kvalifitseerub mudel „mitterahuldavaks”, kuid p-väärtuste alusel võib korrelatsiooni pidada statistiliselt usaldusväärseks ( $p < 0.05$ ). Regressioonsirge tõus on samuti oluliselt erinev soovitud, et see oleks 1. Hea märgina võiks võtta, et sirge on eri jaamades üsna sarnane, mis võib viidata süsteemsele veale mudelis. Sellega saaks arvestada, kui soovitakse mudeli põhjal looduskeskkonna suhtes otsuseid langetada. Determinatsioonikordaja põhjal seletab korrelatsioon parimal juhul ehk Aseri jaamas vaid 47% hajuvusest ning sel juhul on mudeli sooritus „mitterahuldav”. Madalad korrelatsioonikoefitsendid tulevad ilmselt suuremate sügavuste andmetest, kuna joonisel 8 on näha, et mudeli joon varieerub pinnal mõõtmisandmetega hästi.

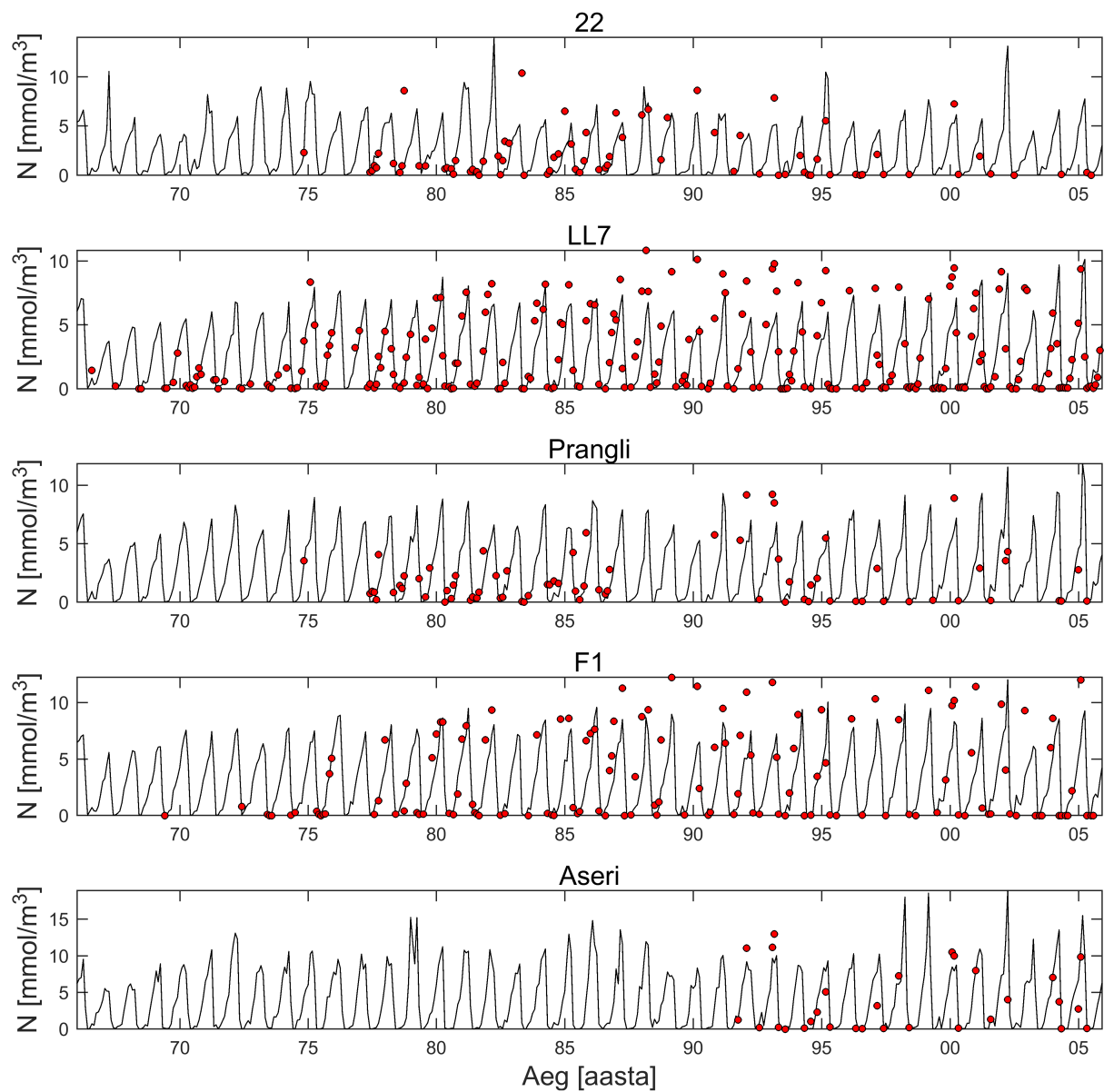
Kaks negatiivset Nash-Sutcliffe'i efektiivsuse väärtust jaamades 22 ja Prangli, mis tulenevad

mõõtmiste standardhälbest suuremast *RMSE* väärtusest, tähendavad, et neis jaamades ennustab mõõtmisandmete keskmine tegelikke väärtusi paremini kui mudel. Jaamas LL7 on tulemus samuti „mitterahuldav”, kuid *E* on positiivne. Jaamades Aseri ja F1 saab mudelit kirjeldada sõnaga „rahuldav”. Kokkuvõttes on mudeli osavus lämmastiku ennustamisel *E* põhjal siiski „mitterahuldav”, olles vaid 0.10.

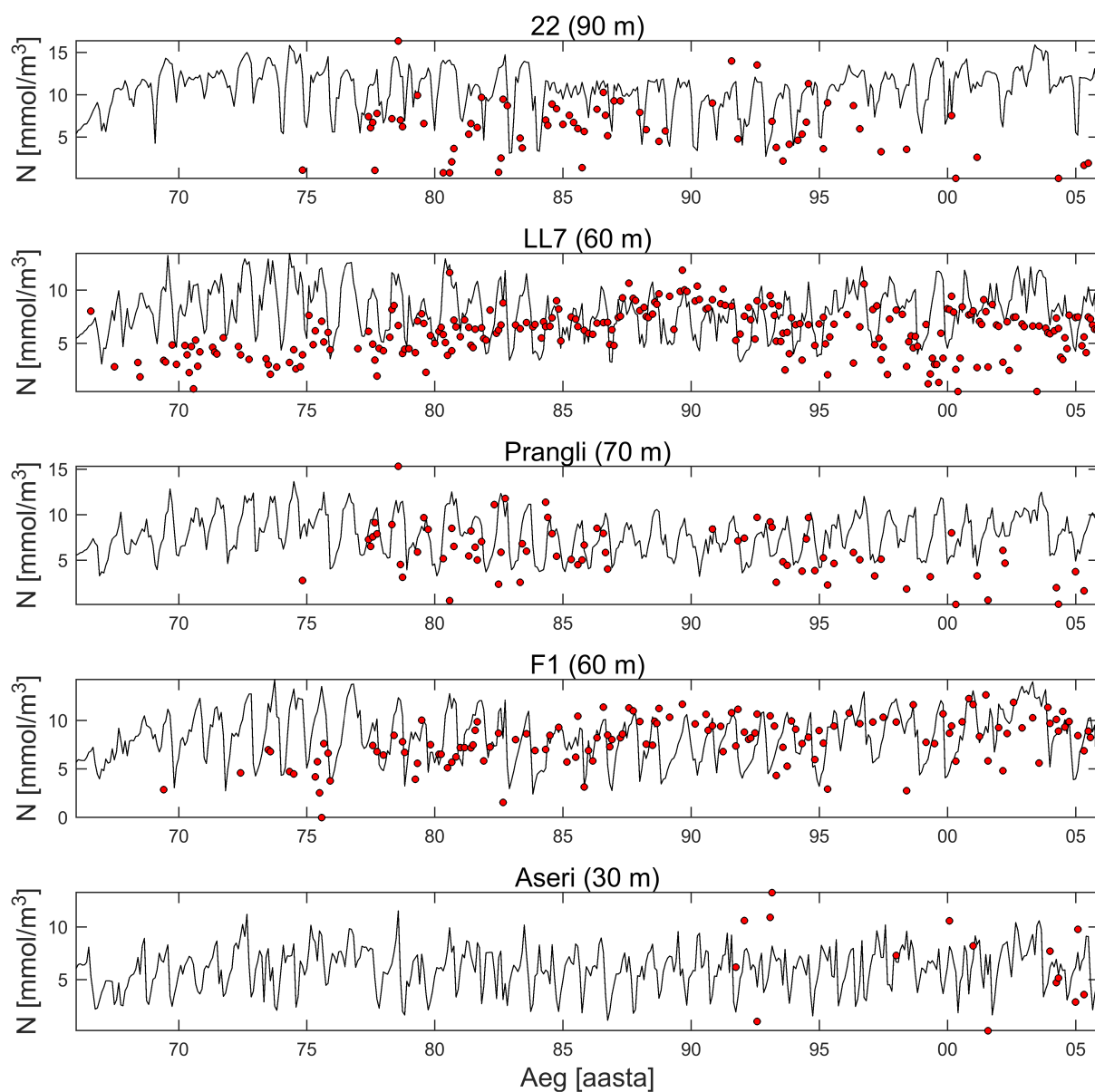
Kulufunktsioon *K* annab taas igas jaamas mudeli osavuseks optimistlikult „väga hea”, olles vahemikus 0.06...0.41.



**Joonis 7.** Lämmastiku hajuvusdiagrammid. Katkendjoon märgib  $y = x$  joont, pidev joon regressioonsirget, mille võrrand on iga graafiku all paremas nurgas.



**Joonis 8.** Lämmastiku aegread pinnal. Joon märgib mudeli andmeid ning punktid mõõtmisandmeid.



**Joonis 9.** Lämmastiku aegread põhjas. Joon märgib mudeli andmeid ning punktid mõõtmisandmeid.

### 3.3 Klorofüll

Tabelis 6 on toodud lämmastikuisaldus fütoplanktonis ning klorofüllis. Nagu mainitud, ei ole teada, missuguse konstandiga korrutada klorofüllis tulemusi, et saada süsinikuisaldust, millele saab Redfieldi suhte alusel vastavusse seada lämmastikuisalduse. Seetõttu ei ole siin toodud statistikuksid  $RMSE$ ,  $E$ ,  $\delta$  ning  $K$ , mille väärtused on mõjutatud mudeli ja mõõtmisandmete erinevast suurusjärgust. Klorofüllis mõõtmisandmed puudusid jaamas 22.

**Tabel 6.** Statistilised näitajad klorofüllil kohta kõikides jaamades, tingimusel, et C:Chl on 1.  $n$  on andmepaaride arv,  $\overline{N_{f\ddot{u}to,m}}$  ja  $\overline{N_{chl,a,v}}$  on vastavalt mudeli andmete ja mõõtmisandmete lämmastiksisalduse keskmine koos standardhälbiga. Viimases tulbas on toodud süsiniku ja klorofüllil keskmise suhe.

| Jaam    | $n$ | $\overline{N_{f\ddot{u}to,m}}$ [mmol/m <sup>3</sup> ] | $\overline{N_{chl,a,v}}$ [mmol/m <sup>3</sup> ] | $N_{chl,a,v}$ vahemik [mmol/m <sup>3</sup> ] | $r$   | $p$    | Keskm. C:Chl |
|---------|-----|---|---|--|-------|--------|--------------|
| 22      | 0   | -   | -   | -  | -     | -      | -            |
| LL7     | 243 | 1.29±1.39   | 0.03±0.04                                       | 0.00...0.18                                  | 0.50  | < 0.05 | 58.25        |
| Prangli | 18  | 0.13±0.23   | 0.04±0.02                                       | 0.02...0.08                                  | 0.63  | < 0.05 | 2.77         |
| F1      | 15  | 0.38±0.65   | 0.03±0.02                                       | 0.00...0.08                                  | -0.12 | 0.68   | 29.20        |
| Aseri   | 3   | 1.21±1.37   | 0.08±0.01                                       | 0.06...0.09                                  | 0.72  | 0.49   | 13.98        |
| Kokku   | 279 | 1.17±1.35   | 0.04±0.03                                       | 0.00...0.18                                  | 0.47  | < 0.05 | 52.64        |

Tabelist 6 on näha, et jaamades LL7 ja Prangli on mudeli fütoplanktoni lämmastiku ning mõõtmiste klorofüllil vahel usaldusväärne korrelatsioon ( $p < 0.05$ ,  $r$  0.50 ja 0.63), kuid tulemused lahterduvad „mitterahuldava” alla. Jaamades F1 ja Aseri ei ole  $p$ -väärtuse ( $p$  0.68 ja 0.49) tõttu mõistlik öelda, et mudel korreleeruks mõõtmisandmetega, kuigi Aseri jaamas  $r = 0.72$ . Väike korrelatsioonikordaja ning suur  $p$ -väärtus neis jaamades on tingitud väikesest mõõtmiste arvust ning asjaolust, et klorofüllis sisalduvat lämmastikku on süsiniku ja klorofüllil ebastabiilse suhte tõttu keeruline hinnata.

Huvi pärast arvutati keskmine C:Chl lokaalselt kui ka üle kõikide andmete. Nii andmete vähesuse kui ka nende ruumilise ja ajalise hajususe tõttu varieerus kordaja jaamati 2.77st 58.25ni, kõiki andmeid vaadates leiti, et klorofüllist süsinikusisalduse arvutamiseks peab klorofüllil andmeid korrutama keskmiselt 52.64ga.

Seejärel korrutati andmeid vastavas jaamas leitud kordajaga ning arvutati statistikud, mis jäeti välja tabelist 6. Tulemused on näha tabelis 8. Võrdluseks on tabelis 7 toodud need, mis esialgsetest tulemustest välja jäeti.

Nagu näha uutest tulemustest tabelis 8, ei ole korrelatsioonikoefitsient muutunud, kuna see kajastab andmepaaride koos varieerumist, mitte nende ar-

**Tabel 7.** Statistikud klorofüllile esialgsete andmetega.

| Jaam    | $RMS E$<br>[mmol/m <sup>3</sup> ] | $E$        | $\delta$ [%] | $K$   |
|---------|-----------------------------------|------------|--------------|-------|
| 22      | -                                 | -          | -            | -     |
| LL7     | 1.86                              | -2794.91   | -3592.58     | 35.68 |
| Prangli | 0.23                              | -124.09    | -246.65      | 4.29  |
| F1      | 0.73                              | -1287.46   | -1216.01     | 16.91 |
| Aseri   | 1.59                              | -20 397.57 | -1419.04     | 83.07 |
| Kokku   | 1.75                              | -2667.81   | -3211.62     | 33.32 |

vulist suhet. Tabelis 7 toodud väärtustega võrreldes on  $E$ ,  $\delta$  ning  $K$  oluliselt paranenud, ent  $E$  põhjal, mis on vahemikus 0.08...-42.15, jääb mudeli sooritus siiski „mitterahuldavaks”.



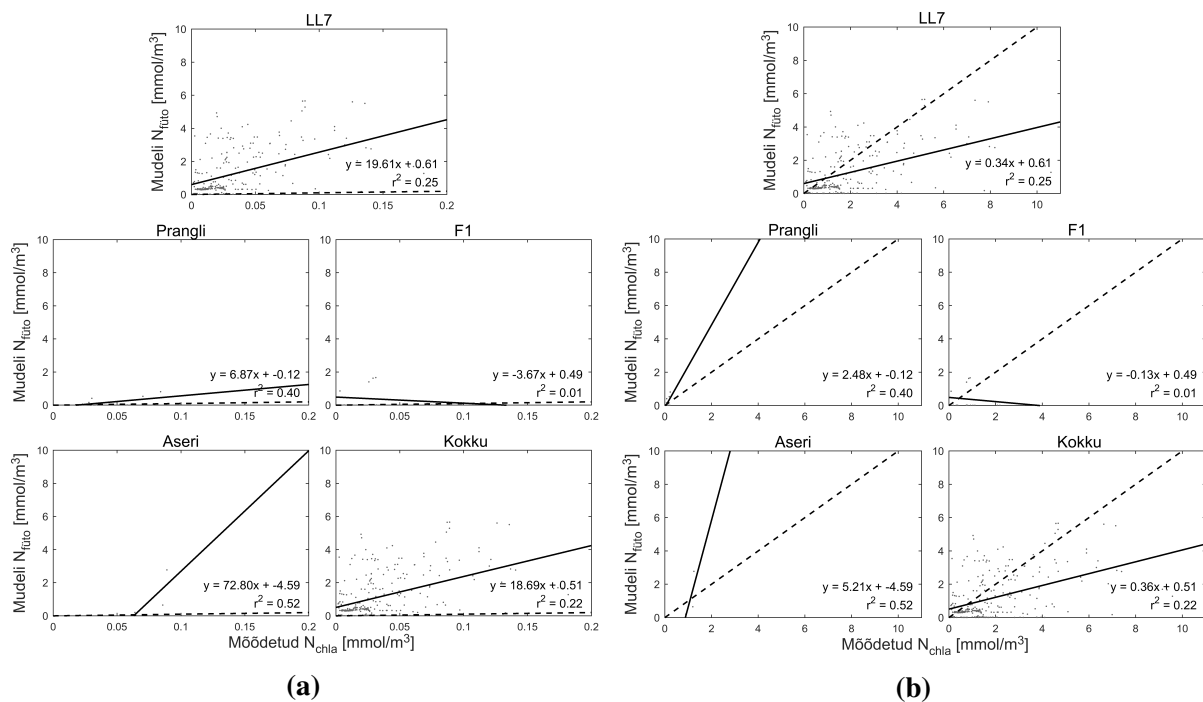
Suhtelise vea järgi võib mudelit „väga heaks” hinnata Aseri jaamas, kus see on  $-8.69\%$ , „heaks” jaamades LL7 ja Prangli ( $|\delta|$  25.15% ja 36.61%) ning „mitterahuldavaks” jaamas F1 ( $\delta = 54.93\%$ ). Kõikide jaamade peale hindab suhteline viga mudelit „rahuldavaks”.

Kulufunktsiooni väärtused on nagu temperatuuri ja lämmastiku korral kõik ühest väiksemad, andes mudeli osavuseks „väga hea”.

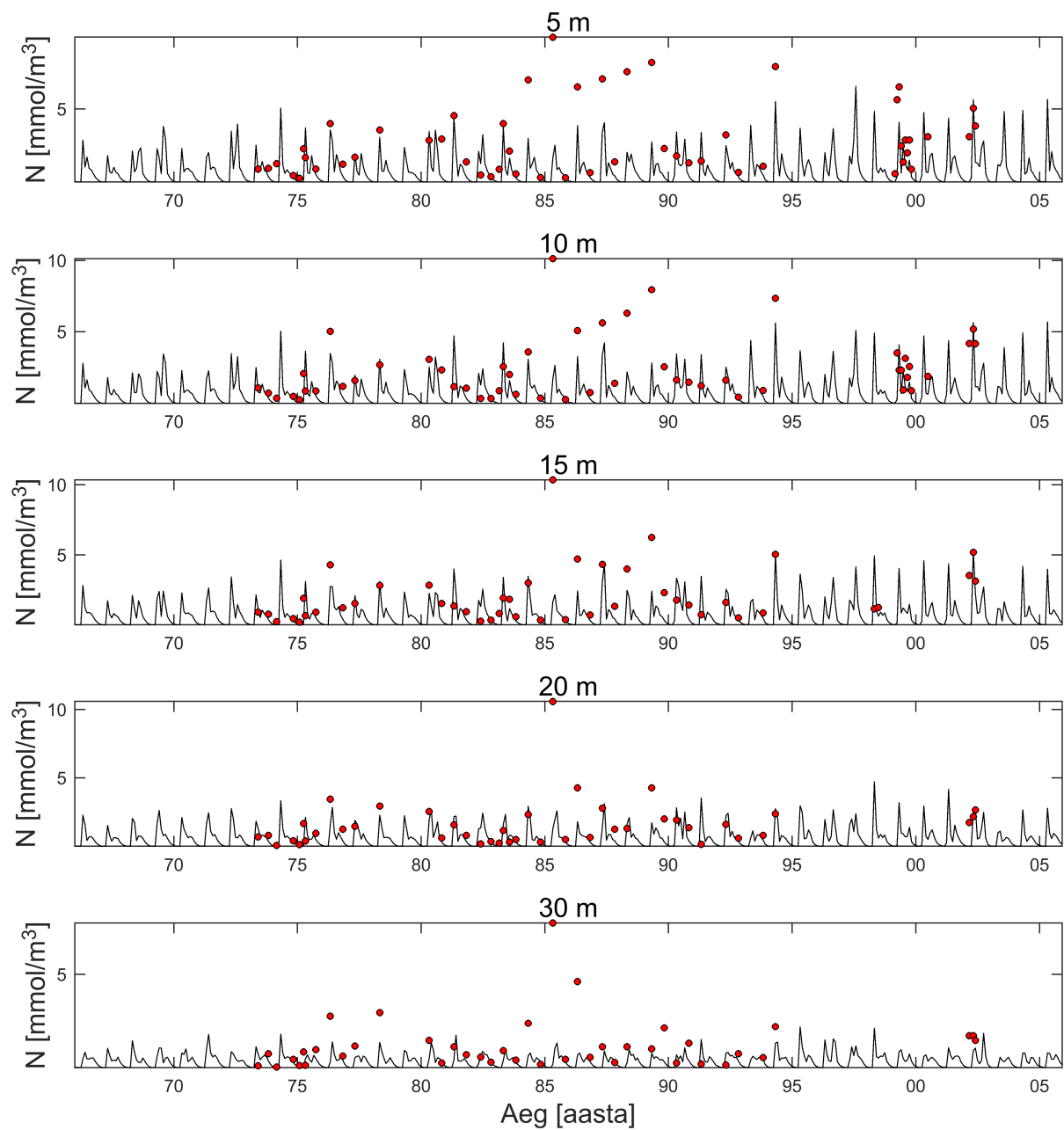
Hajuvusdiagrammidelt joonisel 10 ei tule välja, et mudeli sooritus enne või pärast mõõtmisandmete läbikorrutamist kuigi hea oleks, kuid jaama LL7 aegridadelt joonisel 11 on näha, et mudel läbib mõõtmisandmeid üsna rahuldavalt, alahinnates klorofüllü suuremaid väärtusi suveperioodidel, kuid järgib kenasti aastaringset kõikumist. Teiste jaamade kohta ei kuvatud aegridu väikese mõõtmiste arvu tõttu erinevatel sügavustel. Samal põhjusel kuvati vaid esimesed viis sügavust jaamas LL7.

**Tabel 8.** Statistilised näitajad klorofüllü kohta kõikides jaamades pärast mõõtmisandmete korrutamist keskmise kordajaga (vt tabel 6).  $n$  on andmepaaride arv,  $\overline{N_{\text{füto,m}}}$  ja  $\overline{N_{\text{chla,v}}}$  on vastavalt mudeli andmete fütoplanktoni ja mõõtmisandmete klorofüllü lämmastiksisalduse keskmine koos standardhälbega.

| Jaam    | $n$ | $\overline{N_{\text{füto,m}}}$<br>[mmol/m <sup>3</sup> ] | $\overline{N_{\text{chla,v}}}$<br>[mmol/m <sup>3</sup> ] | $N_{\text{chla,v}}$<br>vahemik<br>[mmol/m <sup>3</sup> ] | $RMSE$<br>[mmol/m <sup>3</sup> ] | $r$   | $p$    | $E$    | $\delta$ [%] | $K$  |
|---------|-----|--|--|--|----------------------------------|-------|--------|--------|--------------|------|
| 22      | 0   | -  | -  | -  | -                                | -     | -      | -      | -            | -    |
| LL7     | 243 | 1.29±1.39  | 2.04±2.05  | 0.03...10.60   | 1.96                             | 0.50  | < 0.05 | 0.08   | 36.61        | 0.36 |
| Prangli | 18  | 0.13±0.23  | 0.10±0.06  | 0.06...0.23  | 0.19                             | 0.63  | < 0.05 | -10.67 | -25.15       | 0.44 |
| F1      | 15  | 0.38±0.65  | 0.85±0.61  | 0.09...2.19  | 1.03                             | -0.12 | 0.68   | -2.02  | 54.93        | 0.76 |
| Aseri   | 3   | 1.21±1.37  | 1.11±0.19  | 0.89...1.24  | 1.02                             | 0.72  | 0.49   | -42.15 | -8.69        | 0.51 |
| Kokku   | 279 | 1.17±1.35  | 1.85±1.79  | 0.03...9.58  | 1.79                             | 0.47  | < 0.05 | -0.01  | 37.08        | 0.38 |



**Joonis 10.** Klorofüllis ja fütoplanktonis sisalduva lämmastiku hajuvusdiagrammid enne kordaja arvestamist (a) ning pärast läbikorrutamist (b). Katkendjoon märgib  $y = x$  joont, pidev joon regressioonsirget, mille võrrand on iga graafiku all paremas nurgas.



**Joonis 11.** Klorofüllis ja fütoplanktonis sisalduva lämmastiku aegread jaamas LL7 pärast mõõtmisandmete korrutamist keskmise C:Chl-ga selles jaamas. Joon märgib mudeli andmeid ning punktid mõõtmisandmeid.

### 3.4 Arutelu

Eelmises peatükis leiti, et mudel ennustab temperatuuri veidi üle, kuid üldjoontes väga hästi: esines tugev korrelatsioon mudeli ja mõõtmisandmete vahel (0.90...0.95), Nash-Sutcliffe'i efektiivsus oli kõrge (0.71...0.88) ning kulufunktsiooni väärtus madal (0.12...0.31). Suhteline viga andis mõnevõrra kehvema hinnangu (−12.57...−25.22%), kuid nii hajuvusdiagrammid koos regressioonsirgega kui ka aegread kinnitasid mudeli väga head sooritust.

Mudel ülehindas ka lämmastikusisaldust, ent osavus jäi kesiseks: korrelatsioonikordaja jäi keskmiseks (0.51...0.68),  $E$  oli ühtlaselt väga halb (−0.39...0.37), suhteline viga varieerus „väga heast” „rahuldavani” (−36.46...11.21%) ning kulufunktsiooni väärtus oli taas madal (0.06...0.41). Ka joonised näitasid mudeli pigem keskmist sooritust. Kuigi pinna lämmastikusisaldust ennustas mudel üsna rahuldavalt, siis põhja väärtusi pigem mitte ning ka hajuvusdiagrammil olid punktid väga laiali.

Pärast mõõtmisandmete läbikorrutamist leitud C:Chl-ga võib mudelit klorofüllil ennustamisel pidada üsna rahuldavaks. Jaamades Prangli, Aseri ja F1 tehtud väheste mõõtmiste tõttu ei ole nende eraldi analüüsil suurt tähtsust. Keskmiseks kordajaks üle kõikide klorofüllil andmete osutus 52.64. Kokku leiti, et läbikorrutatud andmeid ennustab mudel pigem keskmiselt: korrelatsioonikordaja oli tähelepanuväärne, ent üsna madal (0.47),  $E$  oli negatiivne ning  $K$  madal (0.38). Suhteline viga oli päris suur, kuid nii selle kui ka  $E$  negatiivse väärtuse taga peitub ilmselt asjaolu, et süsiniku ja klorofüllil suhe sõltub aastaringselt erinevatest teguritest ning selle täpsem määramine ei kuulunud käesoleva töö juurde. Hajuvusdiagrammid ei jätanud mudelist väga head muljet, kuid jaama LL7 aegridadest paistis, et mudel ennustab varieeruvust üsna hästi, ent alahindab suuremaid väärtusi. Sellest lähtuvalt võib arvata, et C:Chl on madalam kevadsuvel, kui on rohkem klorofüllil ning toimub primaarproduktsoon.

Kasutatud statistikutest tundub kõige asjalikum korrelatsioonikordaja, mis andis sarnase tulemuse silmaga hinnatud aegridadega. Küllaltki hästi sai mudelit hinnata ka Nash-Sutcliffe'i efektiivsuse põhjal. Kehvemaks valikuks osutus kulufunktsioon, mis hindas mudelit igal juhul „väga heaks”.

## Kokkuvõte

Käesoleva töö eesmärgiks oli valideerida ühildatud GETM-ERGOM mudelit Soome lahes, kasutades temperatuuri, nitraadi lämmastikuisalduse ning klorofüllü kontsentratsiooni mõõtmisandmeid viies Soome lahe monitooringujaamas aastatel 1966–2005, viies need vastavusse sama perioodi simulatsiooniandmetega.

Töös leiti, et mudel ennustas hästi temperatuuri, lämmastikuisalduse ning klorofüllü aastaringset varieeruvust. Mudel ülehindas temperatuuri kõikides jaamades, kuid nii statistikuid kui ka graafikuid arvesse võttes oli sooritus temperatuuri ennustamisel nii pinnal kui ka põhjas väga hea. Mudel hindas lämmastikuisaldust pinnal hästi, veidi alahinnates suuremat lämmastikuisaldust talvisel ajal, kuid merepõhjas ei näidanud mudel erilist osavust. Ka statistikute poolest ennustas mudel lämmastikuisaldust pigem rahuldavalt. Kuigi klorofüllü puhul ei olnud võimalik mõõtmis- ja mudeli andmeid usaldusväärselt ühele ühikule teisendada, anti parim, et andmed ligikaudu ühte suurusjärku viia. Selleks tehti kitsendus, et süsinikuisaldus ning klorofüllü kontsentratsioon on üks-ühele, ning leiti lämmastiku kontsentratsioon, kasutades Redfieldi suhet. Saadu baasil leitud keskmise C:Chl põhjal leitud statistikud ning andmete aegread andsid piisavalt infot, et öelda, et mudel ennustab silmnähtavalt ka klorofüllü aastaringset variatsiooni.

Aegridadelt tuli välja, et mudel hindas tegelikkust eri sügavustel erinevalt: mõnda suurst alahindas, mõnda ülehindas. Samuti tuli näiteks lämmastiku aegridadelt välja, et mudel alahindab talvist lämmastikuisaldust pinnal. Seetõttu võiks mudeli osavusest täpsema pildi saamiseks edasistes töodes leida statistikuid erinevatel sügavustel eraldi ning uurida ka seda, kuidas käitub mudel erinevatel aastaaegadel. Samuti oleks tarvis hinnata, mis võiksid olla hindamisskaala ulatused mudelit eri statistikutega valideerides, arvestades just Soome lahe eripärasid. Töö edasiarendusena võiks hinnata mudeli efektiivsust ka teiste väljade nagu soolsuse, fosfaadi ja hapniku simuleerimisel.

Lõputöö eesmärgid saavutati autori hinnangul osaliselt, kuna mudeleid ei saa kunagi täielikult valideerida: leidub mõni statistiline või graafiline meetod, mida pole veel kasutatud, või on tähelepanuta jäänud mõni konkreetne sügavus, aja- või ruumidimensioon.

Võrdlusest saadavad tulemused ei ole võetavad puhta tõena, kuna loodus on pidevas muutumises. Mudel võis teha hea soorituse minevikus, kuid see ei tähenda, et teeb seda ka tulevikus. Seetõttu

peaks mudeli valideerimisest saadud tulemusi vaatama kriitilise pilguga, eriti just seal, kus mudeli ennustusi kasutatakse otsuste tegemiseks. Kõigele vaatamata annab mudeli valideerimine võimaluse võrrelda seda teiste omataolistega, et soovitud keskkonna, parameetri või aja jaoks valida just parimate tulemustega mudel.

## Summary

The aim of this work was to validate coupled GETM-ERGOM model in the Gulf of Finland, which was done by using temperature, nitrogen, and chlorophyll data measured between 1966 and 2005 in five monitoring stations and comparing it with model simulation data from the same time window.

It was found that the skill of the model in terms of simulating temperature, nitrogen and chlorophyll variability was good. Model had the tendency to overestimate temperature in all stations, but considering statistical indices and graphs, it performed very well both on the surface and at the bottom of the sea. On the surface, model underestimated nitrogen concentration in the winter, but was generally good. At the bottom, though, model was not that skilled. Based on statistical metrics, the model was rather satisfying in simulating nitrogen concentration. It was not possible to convert monitoring data of chlorophyll to match the model data of phytoplankton nitrogen due to varying carbon to chlorophyll ratio. Thus, the ratio of 1 was used, hoping the model and monitoring data order of magnitude comes closer. Nitrogen concentration was obtained using Redfield ratio. Resulting data was sufficient to make the conclusion that model is able to simulate the variation of chlorophyll, but the calculated indices were not reliable due to the restriction of C:Chl.

Time series plots indicated that model simulates data differently in terms of different depths, fields and time windows, overestimating the values for one and underestimating for another. Thus, this work could be improved by calculating the statistics for different depths and seasons and also for fields like salinity, phosphorus and oxygen concentration that were not included in this paper. Furthermore, the approximate value ranges of model scores should be estimated for Gulf of Finland.

The author of this work claims to have achieved the purpose of this work only partially because no model is ever fully validated.

The results obtained in this work should not be taken as truth in the sense that nature is in a state of flux. The fact that model may have performed well in the past does not mean it will do the same in the future. The results of model validation should always be treated with criticism especially where decisions are made based on the simulation. Having said that, model validation

gives the opportunity to compare it to other analogous models and choose the one with best scores for desirable environment, parameter or season.



## Viited

- [1] H. Burchard and K. Bolding, “GETM, a general estuarine transport model. Scientific documentation,” tech. rep., European Commission, 2002.
- [2] A. Stips, K. Bolding, T. Pohlmann, and H. Burchard, “Simulating the temporal and spatial dynamics of the North Sea using the new model GETM (general estuarine transport model),” *Ocean Dynamics*, vol. 54, no. 2, p. 266–283, 2004.
- [3] T. Neumann, “Towards a 3d-ecosystem model of the Baltic Sea,” *Journal of Marine Systems*, vol. 25, no. 3–4, p. 405–419, 2000.
- [4] “Nitrogen cycling in the open ocean.” <http://cmore.soest.hawaii.edu/cruises/biolincs/nitrogen.htm>.
- [5] G. Lessin, U. Raudsepp, I. Maljutenko, J. Laanemets, J. Passenko, and A. Jaanus, “Model study on present and future eutrophication and nitrogen fixation in the Gulf of Finland, Baltic Sea,” *Journal of Marine Systems*, vol. 129, p. 76–85, 2014.
- [6] I. Maljutenko and U. Raudsepp, “Validation of GETM model simulated long-term salinity fields in the pathway of saltwater transport in response to the Major Baltic Inflows in the Baltic Sea,” in *2014 IEEE/OES Baltic International Symposium BALTIC*, p. 1–9, May 2014.
- [7] F. Wulff, A. Sokolov, and O. Savchuk, “Nest – a decision support system for management of the Baltic Sea. A user manual,” tech. rep., Baltic Nest Institute, Stockholm University Baltic Sea Centre, August 2013.
- [8] T. Soomere, K. Myrberg, M. Leppäranta, and A. Nekrasov, “The progress in knowledge of physical oceanography of the Gulf of Finland: a review for 1997-2007,” *Oceanologia*, vol. 50, no. 3, p. 287–362, 2008.
- [9] J. J. Cullen, “The deep chlorophyll maximum: Comparing vertical profiles of chlorophyll a,” *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 39, no. 5, p. 791–803, 1982.
- [10] A. C. Redfield, “The biological control of chemical factors in the environment,” *American Scientist*, vol. 46, no. 3, p. 230A–221, 1958.

- [11] A. Andersson and A. Rudehäll, “Proportion of plankton biomass in particulate organic carbon in the northern Baltic Sea,” *Marine Ecology Progress Series*, no. 95, p. 133–139, 1993.
- [12] C. A. Stow, J. Jolliff, D. J. M. Jr., S. C. Doney, J. I. Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead, “Skill assessment for coupled biological/physical models of marine systems,” *Journal of Marine Systems*, vol. 76, no. 1–2, p. 4–15, 2009.
- [13] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics, Springer, 2004.
- [14] E. Jakobson, “Tõenäosusteooria ja statistika II osa – statistika.” 2014.
- [15] D. N. Moriasi, J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, “Model evaluation guidelines for systematic quantification of accuracy in watershed simulations,” *Transactions of the ASABE*, vol. 50, no. 3, p. 885–900, 2007.
- [16] K. Eilola, B. G. Gustafsson, I. Kuznetsov, H. E. M. Meier, T. Neumann, and O. P. Savchuk, “Evaluation of biogeochemical cycles in an ensemble of three state-of-the-art numerical models of the Baltic Sea,” *Journal of Marine Systems*, vol. 88, no. 2, p. 267–284, 2011.
- [17] P. B. Duda, P. R. Humme, A. S. Donigian Jr, and J. S. Imhoff, “BASINS/HSPF: Model use, calibration, and validation,” *Transactions of the ASABE*, vol. 55, no. 4, p. 1523–1547, 2012.
- [18] M. Vichi and S. Masina, “Skill assessment of the PELAGOS global ocean biogeochemistry model over the period 1980–2000,” *Biogeosciences*, vol. 6, p. 2333–2353, 2009.
- [19] D. Marechal, *A soil-based approach to rainfall-runoff modelling in ungauged catchments for England and Wales*. PhD thesis, Cranfield University at Silsoe, 2004.
- [20] J. I. Allen, P. J. Somerfield, and F. J. Gilbert, “Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models,” *Journal of Marine Systems*, vol. 64, no. 1–4, p. 3–14, 2007. Contributions from Advances in Marine Ecosystem Modelling Research, 27–29 June, 2005, Plymouth, {UKAMEMR}.

- [21] M. Maar, E. F. Møller, J. Larsen, K. S. Madsen, Z. Wan, J. She, L. Jonasson, and T. Neumann, “Ecosystem modelling across a salinity gradient from the North Sea to the Baltic Sea,” *Ecological Modelling*, vol. 222, no. 10, p. 1696–1711, 2011.

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Meri Liis Treimann,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

### **Ühildatud GETM-ERGOM mudeli valideerimine Soome lahe temperatuuri, nitraadi ja klorofüllü põhjal aastate 1966–2005 kohta,**

mille juhendaja on Ilja Maljutenko,

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
  3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **26.05.2016**